

TITLE OF THE INVENTION

METHODS FOR DETERMINING NUCLEOTIDE SEQUENCES OF SINGLE NUCLEIC ACID MOLECULES

5

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to a method for determining nucleotide
10 sequences of single nucleic acid molecules by single molecule detection.

Description of the Background

The Sanger method (Proc. Natl. Acad. Sci. USA, 74:5463, 1977) is routinely
used as a method for determining a nucleotide sequence of DNA. This method is also
15 referred to as the dideoxy chain termination method, entailing the steps of annealing a
primer to the 5' end of a DNA sample, synthesizing a complementary chain in the
presence of DNA polymerase and four deoxynucleoside triphosphates (dNTPs where
N denotes A, C, T and G) and each of 2', 3'-dideoxynucleoside triphosphates
(ddNTPs where N denotes A, C, T or G), stopping the elongation reaction at a
20 position where ddNTP is incorporated, and performing gel electrophoresis for the
obtained reaction product, thereby determining a nucleotide sequence of the DNA
sample. The DNA fragments formed at this time are generally labeled with a
radioactive label, which enables the identification of the position of the fragments.

Furthermore, since radioactive labeling requires a special facility, other
25 methods for determining nucleotide sequences which employ fluorescent labels in
place of radioactive labels to detect fluorescence by irradiation of a laser beam, have
also been developed (e.g. Japanese Patent No. 2901004, and JP-B-7-43347).

However, the use of the Sanger method remains problematic as, for example

it requires production of a large number of copies of the DNA to be sequenced by previously incorporating the DNA into a vector, e.g. M13, and that the number of detectable bases is limited to a maximum 1,000 bp per lane of electrophoresis.

Hence, if the nucleotide sequence of a DNA could be determined by directly

5 identifying the bases from the 5' end of the sequence one by one, and an isolated single DNA molecule could be used for direct sequencing without requiring a number of replicated molecules of DNA, efficiency of sequencing could, thereby, be significantly improved.

10 SUMMARY OF THE INVENTION

Accordingly, it is the object of the present invention to provide a method for determining nucleotide sequences of single nucleic acid molecules by single molecule detection. This method advantageous by enables detection of nucleic acid bases one by one, in addition to enabling direct decoding of intracellular DNA or RNA
15 molecules. Furthermore, the present method thereby improves the speed of sequencing by one to two orders of magnitude and enables direct sequencing without producing a number of copies of a single-stranded nucleic acid molecule.

It is, thus, an object of the present invention to provide, in one aspect, a method for determining a nucleotide sequence of a nucleic acid molecule by single dye
20 molecule detection, which entails the steps of:

- (a) immobilizing a nucleic acid molecule onto the surface of a solid;
- (b) annealing a primer, which has a sequence complementary to a part of the sequence of the nucleic acid molecule, to the nucleic acid molecule;
- (c) providing a solution which contains DNA polymerase and one type of
25 dye-labeled dNTP (where N is A, T or U, G or C), or RNA polymerase and one

type of dye-labeled NTP (where N is A, U, G or C), to the immobilized nucleic acid molecule, and allowing the nucleotide to react with the 3' end of the primer, whereby a nucleotide, which forms a base-pair with a base opposed to the reaction site, is bound to the primer by action of the polymerase;

- 5 (d) detecting a bound, dye-labeled dNTP or NTP;
- (e) disrupting the dye molecule of the bound, dye-labeled dNTP or NTP;
- (f) repeating the steps (3) to (5) while changing the type of the dye-labeled dNTP or NTP in turn, to sequentially bind dNTPs or NTPs complementary to the nucleotides of the nucleic acid molecule; and
- 10 (g) determining a nucleotide sequence of the nucleic acid molecule based on the types of the sequentially bound dNTPs or NTPs.

Moreover, it is another aspect, the present invention to provide a method for determining a nucleotide sequence of a nucleic acid by single dye molecule detection, which entails the steps of:

- 15 (a) immobilizing onto the surface of a solid, a primer which has a sequence complementary to a part of the sequence of a nucleic acid molecule to be measured;
- (b) annealing the nucleic acid molecule to the primer;
- (c) providing a solution, which contains DNA polymerase and one type of dye-labeled dNTP (where N is A, T or U, G or C), or RNA polymerase and one type of dye-labeled NTP (where N is A, U, G or C), to the immobilized nucleic acid molecule, and allowing the nucleotide to react with the 3' end of the primer, whereby a nucleotide, which forms a base-pair with a base opposed to the reaction site, is bound to the primer by action of the polymerase;
- 20 (d) detecting a bound, dye-labeled dNTP or NTP;
- 25 (d) detecting a bound, dye-labeled dNTP or NTP;

- (e) disrupting the dye molecule of the bound, dye-labeled dNTP or NTP;
- (f) repeating the steps (3) to (5) while changing the type of the dye-labeled dNTP or NTP in turn, to sequentially bind dNTPs or NTPs complementary to the nucleotides of the nucleic acid molecule; and
- 5 (g) determining a nucleotide sequence of the nucleic acid molecule based on the types of the sequentially bound dNTPs or NTPs.

In one embodiment of the present invention, the solid surface in the above step (1) is the inner wall of a capillary.

In another embodiment of the present invention, the above step (4) comprises
10 optically detecting the dye molecule of the dye-labeled dNTP or NTP. More specifically, the detection can be performed by exciting the dye molecule by irradiation of a laser beam and detecting the thus emitted fluorescence signal. An example of such a detection method is a method, which employs a confocal fluorescence microscope system.

15 In yet another embodiment of the present invention, the disruption of dye molecules in the above step (5) is performed by irradiation of a laser beam stronger than that in step (4).

In still another embodiment of the present invention, the above dye is a fluorescent dye.

20 In a further another embodiment of the present invention, the above solution consists of a droplet, in which an aqueous solution containing the dye-labeled dNTP or NTP, is entrapped within a hydrophobic liquid.

BRIEF DESCRIPTION OF THE DRAWINGS

25 Figure 1 is a schematic diagram showing the method of the present invention

for determining a nucleotide sequence of DNA by single dye molecule detection.

Figure 2 is a schematic diagram showing the procedure to detect a fluorescent signal derived from a dye molecule using a confocal fluorescence microscope system.

5

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention provides a method for determining nucleotide sequences of nucleic acid molecules comprising DNA or RNA by using a single molecule detection method. The single molecule detection method as used herein means a method, in which any signal molecule is detected by any analysis method. In this invention, generally the dye molecule of a dye-labeled deoxynucleoside triphosphate (dNTP, wherein N denotes A, T or U, G or C) or a dye-labeled nucleoside triphosphate (NTP, wherein N denotes A, U, G or C) can be optically detected using a spectroscopic instrument. The detection is performed by using a confocal fluorescence microscope instrument in the Example described below (see Fig. 2).

10
15

The present invention will be now be further described as follows by referring to Fig. 1.

The first step in the method of the present invention entails immobilizing a nucleic acid molecule or a primer having a sequence complementary to a part of the sequence of the nucleic acid molecule, onto the surface of a solid.

20

The nucleic acid molecule can be immobilized onto the solid surface after purification of a nucleic acid sample by standard techniques, for example by preparing single-stranded molecules from the sample by, for example, denaturation with alkaline treatment.

25

The size and type of the primers employed are not specifically limited as long as they can be annealed to nucleic acid molecules. For example, the size of the primers may be at least 10 nucleotides, and normally about 15 to 30 nucleotides.

When a part of the sequence of a nucleic acid molecule to be sequenced is known, a primer may be prepared based on this known sequence and be used. Alternately, random primers or oligo dT primers may be used as the primers.

The solid surface used may be of any material as long as a nucleic acid molecule or a primer can be immobilized thereto. Examples of such a material include glass, quartz and resin. Further, the solid surface may be flat, curved or in any other form. For example, the inner wall of a capillary (e.g. made of glass, quartz or resin) can be used as a solid surface. A capillary is appropriate for automatically injecting a solution containing dye-labeled dNTP or NTP and polymerase enzyme into the inside of the capillary after immobilization of nucleic acid molecules or primers within the capillary. The internal diameter of a capillary is, for example, approximately 100 to 250 μ m, and the sufficient length is generally about 10 to 50mm, but the internal diameter and length are not limited thereto.

Preferably, the solid surface is previously treated so as to facilitate immobilization of a nucleic acid molecule to be sequenced or a primer, and so as not to allow unreacting dye-labeled dNTP or NTP molecules to adsorb to the surface.

Binding of a nucleic acid molecule or a primer to the solid surface can be performed by, for example, a general UV crosslinking method. More specifically, a solution prepared by dissolving a nucleic acid in a Carnoy's solution (methanol/acetic acid (3:1 v/v)) is introduced into a quartz glass capillary, then dried and solidified at room temperature. Next, 2xSSC (NaCl 1.75g, sodium citrate 0.882g/100ml) is introduced into the capillary, followed by UV irradiation, thereby binding the nucleic acid to the

solid surface. It is preferable to immobilize one molecule of nucleic acid onto the solid surface, but in practice, multiple nucleic acid molecules are immobilized onto the surface. In an actual measurement, diluted nucleic acid molecules (approximately 0.1 to 100pmol/ μ l, preferably 30 to 70pmol/ μ l) are immobilized onto the solid surface, one immobilized nucleic acid molecule enters into the field of vision, and then sequencing is performed.

The second step entails annealing a primer to the nucleic acid molecule on the solid surface, or annealing the nucleic acid molecule to a primer on the solid surface.

The third step entails providing a solution containing DNA polymerase and one type of dye-labeled dNTP (where N is A, T or U, G or C) or containing RNA polymerase and one type of dye-labeled NTP (where N is A, U, G or C) to the immobilized nucleic acid molecule, and allowing a nucleotide to react with the 3' end of the above primer. At this time, a nucleotide, which forms a base-pair with a base opposed to the reaction site, is bound to the primer by action of the polymerase.

As used herein, the term "one type of" means a certain one of four types of dye-labeled dNTP or NTP. As the result that this "one type" is specified, the type of dNTP or NTP to be actually bound to a primer would be known.

In Fig. 1, a solvent is allowed to flow through a capillary cell in which a nucleic acid molecule has been immobilized, and a solution containing DNA polymerase and only one type of dye-labeled base dATP is introduced into the flow, in order to cause the reaction of the base next to a DNA sequence ATG previously annealed to a target DNA. If no incorporation reaction of the base by DNA polymerase occurs, no dye is detected on the DNA by the single molecule detection because unreacting bases are washed away. In this case, the same procedure is

performed for a solution containing another type of base (dGTP, dCTP or dTTP).

This procedure is repeated until a base is incorporated into the above DNA sequence and a dye is detected on the DNA. As shown in the figure, which nucleotide follows ATG (in this case, T follows ATG) will be known by allowing a solution containing a dye-labeled dTTP to contact with the target DNA which has been annealed to a primer with ATG already bound to the 3' end, and then detecting the dye. For a DNA sample having the same sequential bases, multiple bases may be incorporated into a DNA sequence. In such a case, identification of the number of dyes by, for example, fluorescence intensity, enables detection of the number of sequential bases incorporated.

In the present invention, a solution containing DNA polymerase and one type of dye-labeled dNTP (where N is A, T or U, G or C), or RNA polymerase and one type of dye-labeled NTP (where N is A, U, G or C), can consist of a droplet in which an aqueous solution which contains the dye-labeled dNTP or NTP is entrapped within a hydrophobic liquid such as mineral oil. Such a droplet can be easily prepared using, for example, a micro-injector. The size of a droplet is, for example, approximately 10 to 25 μ m in diameter, corresponding to several hundreds fL in volume.

Examples of a solvent to flow through a capillary cell are those in which dye-labeled dNTP or NTP and polymerase can be dissolved, including a buffer containing, for example, 67mM KPO₄ (pH7.5), 6.7mM MgCl₂, and 1mM 2-mercaptoethanol. When a droplet is used, a preferred solvent is one that has no affinity with the droplet. For example, light white oil (d=0.84g/ml; general commercial name: mineral oil) can be used.

Examples of dyes for labeling dNTP or NTP include fluorophors or luminophors, such as rhodamine and fluorescein (e.g. tetramethyl rhodamine, TMR,

emission wavelength: 570nm; tetramethyl rhodamine isothiocyanate, TRITC, emission wavelength: 573nm; Rhodamine 6G, emission wavelength: 550nm; fluorescein isothiocyanate, FITC, emission wavelength: 515nm). In addition, 4-fluoro-7-nitro-benzofurazon (NBD-F, emission wavelength: 540nm), Texas red (emission wavelength: 605nm) or the like can be used. In the method of this invention, the same dye may be used regardless of the type of bases, that is dNTP or NTP, or the dye may differ according to the type of bases. To simplify the procedure, dNTP or NTP is preferably labeled with the same dye. In binding a dye to dNTP or NTP, a commercially available product (i.e., dye-labeled dNTP or NTP) may be used, or otherwise, such a product may be synthesized according to a method described in literature (e.g. *J. Histochem. Cytochem.* 44(5):525-529, 1996).

The fourth step entails detecting a bound, dye-labeled dNTP or NTP.

As shown in Fig. 2, a bound, dye-labeled dNTP or NTP is detected by irradiation of a laser beam to the nucleic acid molecule using, for example, a confocal fluorescence microscope system, and introducing fluorescent signals emitted from the excited dye molecule into a detector to count the number of photons and thereby detect the fluorescent signals.

In Fig. 2, the excitation light (488nm) of an argon ion laser is reflected by a dichroic mirror to focus on a DNA sample through an objective lens, a fluorescent signal emitted from the dye molecules excited by the excitation light is introduced into a confocal pin hole (e.g. 50 μ m in diameter) through a band pass filter, and then the number of photons that have reached a detector (e.g. avalanche photodiode) is counted by a multichannel counter, thereby detecting the fluorescent signal. The presence of a band pass filter enables to selectively take fluorescent signals. Further the presence of a pin hole eliminates unnecessary light.

The fifth step entails disrupting dye molecules of bound, dye-labeled dNTP or NTP.

The method of the present invention requires disruption of dye molecules, after a dye molecule is detected after reaction of dye-labeled dNTP or NTP on a nucleic acid molecule, but before the next dye-labeled dNTP or NTP is bound. As a means for this purpose, for example a method of irradiation of a laser beam, which is stronger than that in step (4) (e.g., about 10mW) can be used.

The sixth step entails sequentially binding a dNTP or NTP, which forms a base-pair with the nucleotide of the nucleic acid molecule, by repeating the above steps (3) to (5) while changing the type of dye-labeled dNTP or NTP in turn.

As used herein, the word "changing the type of dye-labeled dNTP or NTP in turn" means to keep changing the type of base introduced until binding occurs. That is, if no binding occurs when a certain base of the four types of dye-labeled dNTP or NTP is delivered at a nucleic acid molecule, then another certain base is delivered. If no binding occurs again, then yet another certain base is delivered at the nucleic acid molecule. Whether or not binding of bases occurs is confirmed by step (4), and the dye molecule of a bound base is disrupted in step (5). The procedures from steps (3) to (5) are repeated sequentially up to the (maximum) number of bases of the nucleic acid molecule.

The seventh step entails determining a nucleotide sequence of the nucleic acid molecule based on the types of the sequentially bound dNTPs or NTPs.

EXAMPLES

The present invention will be now described in more detail by reference to certain Examples which are provided solely for illustration and are not intended to be

limitative.

Example 1

A Capton-coated quartz glass capillary (200 μ m in internal diameter x 20mm in length; purchased from GL Science, Japan) was heated with a burner, thereby burning and removing a portion of the Capton coating to provide an observation window for a microscope. The capillary was immersed in 1M KOH, and then in a conc. H₂SO₄/30%H₂O₂ (1:2v/v) mixture, thereby eliminating oil or organic matter attached on the glass surface and washing. Next, a template DNA was immobilized onto the inner wall of the glass capillary by the UV crosslinking method. That is, a solution prepared by dissolving DNA in a Carnoy's solution (methanol/acetic acid (3:1 v/v)) was introduced into the inside of the glass capillary, then dried and solidified at room temperature. Subsequently, 2 x SSC (NaCl 1.75g and sodium citrate 0.882g/100ml) was introduced into the capillary, followed by irradiation with UV, thereby binding DNA to the solid surface. Here, the concentration of the template DNA in the Carnoy's solution was 50pmol/ μ l, so that DNA could be immobilized at a density sparse enough to confirm the reaction on a single template DNA.

Next, DNA polymerase reaction was performed for the template DNA. The reaction solution was prepared by dissolving a primer having a sequence complementary to the template DNA, DNA polymerase (Klenow Fragment, purchased from TOYOBO) and a dye-labeled nucleotide in a buffer solution (67mM KPO₄, pH7.5, 6.7mM MgCl₂, 1mM 2-mercaptoethanol). The reaction solution was allowed to react with the template DNA, performing an incorporation reaction.

After reaction, the reaction solution was removed, and then the same solvent as described above was allowed to flow through the capillary for washing. Then, the

inner wall of the glass capillary was observed for a bound dye using a confocal fluorescence microscope system. That is, the dye molecule of a dye-labeled nucleotide was excited by a laser beam, and the resulting fluorescence was observed. At this time, the template DNA was previously immobilized on the inner wall surface of the glass capillary at a density sparse enough to allow confirmation of an incorporation reaction on a single template DNA.

More detailed description of experimental examples will be given, as follows.

10 Reagents used:

Template DNA No. 1 (SEQ ID NO: 1)

5'-CTG CTC ATA TAT ATA TAG GTG CCA GTC GGA TAG TGT T-3'

Template DNA No. 2 (SEQ ID NO: 2)

5'-GCG GAG GAA GGT CCT TGG TCA TTA GGA TCC-3'

15 Primer No. 1 (SEQ ID NO: 3)

5'-AAC ACT ATC CGA CTG GCA CC-3'

Primer No. 2 (SEQ ID NO: 4)

5'-GGA TCC TAA TGA CCA AGG-3'

Dye-labeled nucleotides

20 BODYPY-TMR-dUTP (purchased from FUNAKOSHI ;
absorbance wavelength 544nm, fluorescence wavelength 570nm)

TMR-dATP (purchased from Daiichi Pure Chemicals ;
absorbance wavelength 550nm, fluorescence wavelength 570nm)

25 TMR-dGTP (purchased from Daiichi Pure Chemicals ; absorbance wavelength
550nm, fluorescence wavelength 570nm)

Experiment 1 and results:

BODYPY-TMR-dUTP nucleotide was allowed to react with a combination of the template DNA No. 1 and the primer No. 1. In this combination of DNAs, the template DNA has a sequence such that dUTP is incorporated onto the template DNA. The experiment was performed according to the above described method, and as a result, it was confirmed by observing the presence of fluorescence from a single fluorescence molecule that BODYPY-TMR-dUTP nucleotide was incorporated onto the single template DNA. Further, reference test 1 was conducted to confirm that the observed fluorescence was not due to the BODYPY-TMR-dUTP nucleotide non-specifically attached to the inner wall of a glass capillary. That is, the above reaction solution was allowed to react with the surface of a glass capillary containing no template DNA, and as a result, it was confirmed that BODYPY-TMR-dUTP nucleotide was not attached to the glass capillary surface because no fluorescence was observed. Moreover, reference test 2 was conducted using TMR-dATP nucleotide instead of BODYPY-TMR-dUTP nucleotide. In this case, it is predicted that TMR-dATP would not be incorporated into the single template DNA. Actually, it was confirmed that the nucleotide was not incorporated onto the single template DNA because no fluorescence was observed.

Experiment 2 and results:

The same experiment was repeated with the exception that a combination of the template DNA No. 2 and the primer No. 2 was used. In this combination of DNAs, the template DNA has a sequence such that dATP is incorporated onto the template DNA. Comparison of the reaction of TMR-dATP nucleotide to that of

TMR-dGTP nucleotide revealed that TMR-dATP was incorporated but TMR-dGTP was not incorporated onto the single template DNA.

Experiment 3 and results:

5 A laser (10 mW, 488 nm) irradiates, for several seconds, the sample from Experiment 1, in which BODYPY-TMR-dUTP has been incorporated by using a combination of the single template DNA No. 1 and the primer No. 1 on the inner wall of the capillary, thereby disrupting the dye molecule of the dye-labeled nucleotide. Similarly, the sample from Experiment 2, in which TMR-dATP has been incorporated
10 by using a combination of the single template DNA No. 2 and the primer No. 2 on the inner wall of the capillary was treated in the same manner as that described above, so that the dye molecule of the dye-labeled nucleotide was disrupted. For these samples, the reaction of dye-labeled nucleotides was subsequently repeated so that the types of nucleotides incorporated could be specified one after another.

15 As may readily be appreciated from the above, the present invention enables the decoding of the bases in nucleotide sequences of a nucleic acid, one by one.

 The present invention also provides a kit which may be used in practicing the present invention. The kit may generally contain at least i) one or more solutions each containing DNA polymerase and one type of dye-labeled dNTP and/or ii) one or more
20 solutions each containing RNA polymerase and one type of dye-labeled NTP.

 Optionally, other constituents may be included therein, such as an immobilizing solid surface, such as glass, quartz or resin in any form such as flat, curved or even a capillary, which is preferably pre-treated so as to facilitate immobilization of a nucleic acid thereto.

25 Further, the kit may generally contain one or a multiplicity of types of dye-

labeled dNTP or dye-labeled NTP solutions. The solutions may be an aqueous solution-containing the labeled dyes entrapped in a hydrophobic liquid, such as mineral oil.

5 Additionally, the kit may be used in conjunction with a laser for irradiation and a confocal fluorescence microscope system for detection, for example, or the kit may include such items.

SEQUENCE LISTING FREE TEXT

10 SEQ ID NO:1 – Description of Artificial Sequence: synthesized template DNA
SEQ ID NO:2 – Description of Artificial Sequence: synthesized template DNA
SEQ ID NO:3 – Description of Artificial Sequence: a primer
SEQ ID NO:4 – Description of Artificial Sequence: a primer